

A simplified approach to parameter estimation and selection of sparse, mean reverting portfolios

Norbert Fogarasi / János Levendovszky

Received 2011-11-24, accepted 2012-09-22

Abstract

In this paper, we study the problem of finding sparse, mean reverting portfolios in multivariate time series. This can be applied to developing profitable convergence trading strategies by identifying portfolios which can be traded advantageously when their prices differ from their identified long-term mean. Assuming that the underlying assets follow a VAR(1) process, we propose simplified, dense parameter estimation techniques which also provide a goodness of model fit measure based on historical data. Using these dense estimated parameters, we describe an exhaustive method to select an optimal sparse mean-reverting portfolio which can be used as a benchmark to evaluate faster, heuristic methods such as greedy search. We also present a simple and very fast heuristic to solve the same problem, based on eigenvector truncation. We observe that convergence trading using these portfolio selection methods is able to generate profits on historical financial time series.

Keywords

mean reversion · sparse estimation · convergence trading · parameter estimation · VAR(1) model · covariance selection · financial time series

1 Introduction

Mean reversion, as a classic indicator of predictability in financial markets, has received a lot of attention over the last few decades. It has been shown that equity excess returns over long horizons are mean-reverting and therefore contain an element of predictability [8, 10, 13]. Convergence trading, by estimating the parameters of mean reverting portfolios has also been proposed and studied in a number of previous research publications [2, 7]

In his recently published article, d'Aspremont in [5] posed the problem of finding mean-reverting portfolios which are sparse. While there exist simple and reliable methods to identify mean reversion in univariate time series, selecting portfolios from multivariate data which exhibit this property is a much more difficult problem. This can be approached by the Box-Tiao procedure [3] to extract cointegrated vectors by solving a generalized eigenvalue problem. On the other hand, sparseness, he argues, is desirable for reducing transaction costs associated with convergence trading as well as for increasing the interpretability of the resulting portfolio. He developed a new approach to solve the problem by using semidefinite relaxation and compared the efficiency of this solution to the simple greedy algorithm in a number of markets.

In this paper, we further extend their work by developing a new approach to identifying parameters of the stationary first order vector autoregressive VAR(1) model and we propose a new benchmark for evaluating their proposed polynomial time heuristics to this NP hard problem. Our proposed benchmark is the computationally expensive, but optimal, exhaustive solution. This is an important step missing in [5] which shows the overall reliability and performance of the investigated methods.

The structure of the paper is as follows.

- In Section 2, after giving a formal presentation of the problem and we explain how the optimal solution can be found by exhaustive search and introduce the truncation method as a very fast alternative to greedy search.
- In section 3, a simplified approach for the estimation of the model parameters of a VAR(1) model is discussed and a new measure is also developed which shows the goodness of fit of

Norbert Fogarasi

Department of Networked Systems and Services, BME, H-1117 Budapest,
Magyar Tudósok körútja 2., Hungary
e-mail: fogarasi@hit.bme.hu

János Levendovszky

Department of Networked Systems and Services, BME, H-1117 Budapest,
Magyar Tudósok körútja 2., Hungary
e-mail: levendov@hit.bme.hu

the data to the model.

- In section 4, the methodology on generated VAR(1) data is validated and significant trading gains are demonstrated on historical time series of real data, the daily close prices of stocks comprising the S&P 500 index.
- Finally, in section 5 some conclusions are drawn and directions for future research are outlined.

2 Sparse mean reverting portfolio selection

In this section the model is described together with the foundations of identifying mean reverting portfolios. Our approach follows the one published in [5], however, in section 2.3 we develop novel heuristic approaches to fulfill the cardinality constraint.

2.1 Mean reverting portfolios

Low frequency trading is based on identifying mean reverting portfolios which follow the so-called Ornstein-Uhlenbeck process [12]. This process is characterized by the following stochastic differential equation

$$dp(t) = \lambda(\mu - p(t))dt + \sigma dW(t) \quad (1)$$

where $W(t)$ is a Wiener process and λ (mean reversion coefficient), μ (long-term mean) and σ (volatility) are constants. By using the Ito-Doeblin formula [9], one can obtain the following solution:

$$p(t) = p(0)e^{-\lambda t} + \mu(1 - e^{-\lambda t}) + \int_0^t \sigma e^{-\lambda(t-s)} dW(s) \quad (2)$$

which implies that

$$E[p(t)] = p(0)e^{-\lambda t} + \mu(1 - e^{-\lambda t}) \quad (3)$$

and asymptotically

$$\lim_{t \rightarrow \infty} p(t) \sim N\left(\mu, \sqrt{\frac{\sigma^2}{2\lambda}}\right) \quad (4)$$

For trading, λ is a key parameter, as it determines how fast the process gets back to the mean, as well as inversely indicating the level of uncertainty around the mean (via the standard deviation of the asymptotic Gaussian distribution). Hence, the larger the λ , the more suitable is the mean reverting portfolio for convergence trading, as it quickly returns to the mean and it contains a minimum amount of uncertainty around the mean. Therefore, we will be concerned with finding sparse portfolios which are optimal in the sense that they maximize λ .

2.2 Mean reverting portfolio as a generalized eigenvalue problem

In this section we view the asset prices as a stationary, first order, vector autoregressive VAR(1) process. Let $s_{i,t}$ denote the price of asset i at time instant t , where $i = 1, \dots, n$ and t are

positive integers and assume that $\mathbf{s}_t^T = (s_{1,t}, \dots, s_{n,t})$ is subject to a first order vector autoregressive process, VAR(1), defined as follows:

$$\mathbf{s}_t = \mathbf{A}\mathbf{s}_{t-1} + \mathbf{W}_t, \quad (5)$$

where \mathbf{A} is an $n \times n$ matrix and $\mathbf{W}_t \sim N(0, \sigma I)$ are i.i.d. noise terms for some $\sigma > 0$

One can introduce a portfolio vector $\mathbf{x}^T = (x_1, \dots, x_n)$, where component x_i denotes the amount of asset i held. In practice, assets are traded in discrete units, so $x_i \in \{0, 1, 2, \dots\}$ but for the purposes of our analysis we allow x_i to be any real number, including negative ones which denote the ability to short sell assets. We seek the optimal portfolio vector exhibiting mean reverting property under sparseness constraint, i.e. $\text{card}(\mathbf{x}) \leq L$ where card denotes the number of non-zero components and L is a given positive integer $1 \leq L \leq n$. Multiplying both sides with vector \mathbf{x} (in the inner product sense), we obtain

$$\mathbf{s}_t^T \mathbf{x} = \mathbf{s}_{t-1}^T \mathbf{A} \mathbf{x} + \mathbf{W}_t^T \mathbf{x} \quad (6)$$

Defining the predictability factor

$$v(\mathbf{x}) := \frac{E(\mathbf{x}^T \mathbf{s}_{t-1}^T \mathbf{A}^T \mathbf{s}_{t-1} \mathbf{x})}{E(\mathbf{x}^T \mathbf{s}_t \mathbf{s}_t^T \mathbf{x})}, \quad (7)$$

which we can take as a proxy for the portfolio's mean reversion. Maximizing this expression will yield the following optimization problem for finding the best portfolio vector \mathbf{x}_{opt}

$$\mathbf{x}_{opt} : \max_{\mathbf{x}} v(\mathbf{x}) \sim \max_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{G} \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}} \quad (8)$$

under the constraint $\text{card}(\mathbf{x}) \leq L$, where \mathbf{G} is the stationary covariance matrix of the \mathbf{s}_t process. Based on (8) we can see that the problem is equivalent to finding the eigenvector corresponding to the maximum eigenvalue in the following generalized eigenvalue problem [5]:

$$\mathbf{A}^T \mathbf{G} \mathbf{A} \mathbf{x} = \lambda \mathbf{G} \mathbf{x} \quad (9)$$

which can then be solved as

$$\det(\mathbf{A}^T \mathbf{G} \mathbf{A} - \lambda \mathbf{G}) = 0 \quad (10)$$

under the cardinality constraint. Note that this can be transformed into a traditional eigenvalue problem by introducing the variable $\mathbf{u} := \mathbf{G}^{1/2} \mathbf{x}$ so that we have

$$\mathbf{G}^{-1/2} \mathbf{A}^T \mathbf{G} \mathbf{A} \mathbf{G}^{-1/2} \mathbf{u} = \lambda \mathbf{u} \quad (11)$$

where the cardinality constraint is now placed upon $\mathbf{G}^{-1/2} \mathbf{u}$.

2.3 Sparse portfolio selection

In the previous section we have outlined how to select a portfolio which maximizes predictability by solving a generalized eigenvalue problem. However, the cardinality constraint poses an additional computational challenge as the number of subspaces into which the optimality must be checked grows exponentially. In fact, Natarajan shows that this problem is equivalent

to the subset selection problem which is proven to be NP-hard [11]. However, as a benchmark metric, we can compute the optimal solution which, depending on the level of sparsity and the total number of assets, could be computationally feasible. We also describe a polynomial time heuristic algorithm for an approximate solution of this problem.

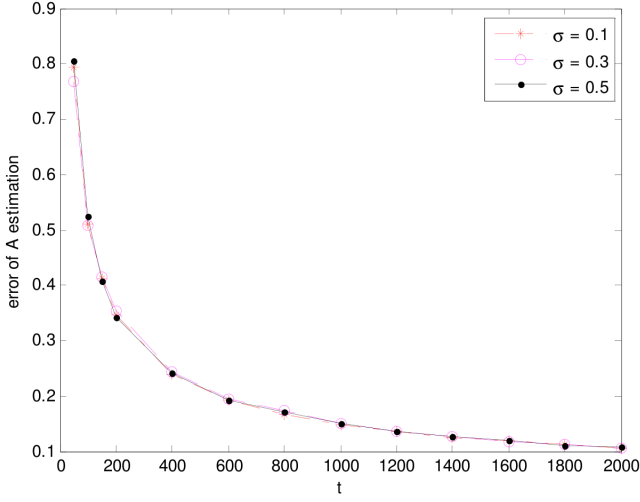


Fig. 1. $\|\hat{\mathbf{A}} - \mathbf{A}\|$ vs. t for $n=8$, $\sigma=0.1, 0.3, 0.5$, generating 100 independent time series for each t and plotting the average norm of error

2.3.1 Exhaustive search method

The brute force approach of constructing all $\binom{n!}{L(n-L)t}$ L -dimensional submatrices of \mathbf{G} and $\mathbf{A}^T \mathbf{G} \mathbf{A}$ and then solving all the corresponding eigenvalue problems to find the theoretical optimum is, in general, computationally infeasible. However, for relatively small values of n and L , or as a benchmark computed off-line, this method can provide a very useful basis of comparison. Indeed, for the practical applications considered in [5] (selecting sparse portfolios of $n=8$ U.S. swap rates and $n=14$ FX rates), this method is fully applicable and can be used to see the level of sub-optimality of other proposed methods.

2.3.2 Greedy method

A reasonably fast heuristic, first presented by d'Aspermont in [5] is the so-called greedy method which we will briefly explain. Let I_k the set of indices belonging to the k (or smaller number) non-zero components of \mathbf{x} . One can then develop following recursion for constructing I_k with respect to k .

When $k = 1$, we simply find

$$l_1 = \arg \max_{i \in [1, n]} \frac{(\mathbf{A}^T \mathbf{G} \mathbf{A})_{ii}}{\mathbf{G}_{ii}}$$

, $i \in [1, n]$. Suppose now that we have a good approximate solution with support set I_k given by

$$(\mathbf{x})_k = \arg \max_{\mathbf{x} \in \mathbb{R}^n: \mathbf{x}_{I_k^c} = 0} \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{G} \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}}$$

where I_k^c is the complement of the set I_k . This can be solved as a generalized eigenvalue problem of size k . We seek to add one variable with index i_{k+1} to the set I_k to produce the largest increase in predictability by scanning each of the remaining indices in I_k^c . The index i_{k+1} is then given by

$$i_{k+1} = \arg \max_{i \in I_k^c} \max_{\mathbf{x} \in \mathbb{R}^n: \mathbf{x}_{J_i} = 0} \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{G} \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}} \quad \text{where } J_i = I_k \cup \{i\} \quad (12)$$

which amounts to solving $(n - k)$ generalized eigenvalue problems of size $k + 1$. We then define $I_{k+1} = I_k \cup \{i_{k+1}\}$, and repeat the procedure until $k = n$. Naturally, the optimal solutions of the problem might not have increasing support sets $I_k \subset I_{k+1}$, hence the solutions found by this recursive algorithm are potentially far from optimal. However, the cost of this method is relatively low: with each iteration costing $O(k^2(n - k))$, the complexity of computing solutions for all target cardinalities k is $O(n^4)$. This recursive procedure can also be repeated forward and backward to improve the quality of the solution.

2.3.3 Truncation method

A simple and very fast heuristic that we can apply is the following. First, compute \mathbf{x}_{opt} , the unconstrained, n -dimensional solution of the optimization problem in (8) by solving the generalized eigenvalue problem in (9). Next, consider the L largest values of \mathbf{x}_{opt} and construct $L \times L$ dimensional submatrices \mathbf{G}' and $(\mathbf{A}^T \mathbf{G} \mathbf{A})'$ corresponding to the L largest dimensions. Solving the generalized eigenvalue problem in this reduced space and padding the resulting \mathbf{x}'_{opt} with 0's will yield a feasible solution \mathbf{x}_{opt}^{trunc} to the original constrained optimization problem. The big advantage of this method is that with just two maximum eigenvector computations, we can determine an estimate for the optimal solution. The intuition behind this heuristic is that the heaviest dimensions in the solution of the unconstrained optimization problem could provide, in most cases, a reasonable guess for the dimensions of the constrained problem. This is clearly not the case in general, but nonetheless, the truncation method has proven to be a very quick and useful benchmark for evaluating other methods.

3 Estimation of model parameters

As explained in the preceding sections, in the knowledge of the parameters \mathbf{G} and \mathbf{A} , we can apply various heuristics to approximate the L -dimensional optimal sparse mean-reverting portfolio. However, these matrices must be estimated from the historical observations of the random process \mathbf{s}_t . There is a vast literature on the topic of parameter estimation of VAR(1) processes, recent research has focused on sparse and regularized covariance estimation [1, 4, 14]. However, our approach will be to gain a dense estimate for \mathbf{G} which best describes the observed historical data and to deal with dimensionality reduction with the more sophisticated apparatus outlined in Section 2. Another

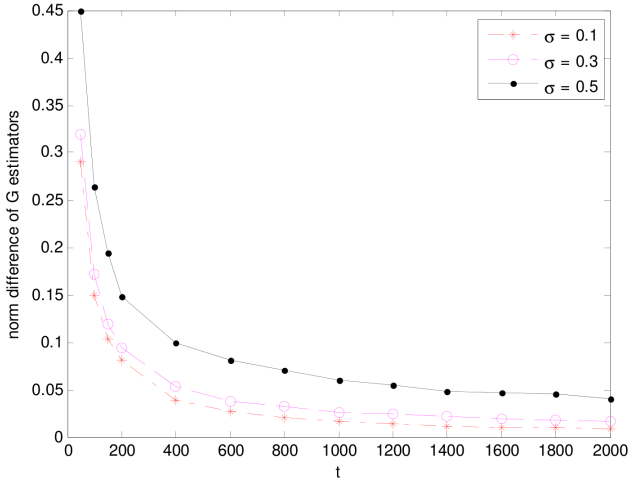


Fig. 2. $\|\hat{\mathbf{G}}_1 - \hat{\mathbf{G}}_2\|$ vs. t for $n=8$, $\sigma=0.1, 0.3, 0.5$, generating 100 independent time series for each t and plotting the average norm of error

important objective that we pose for the parameter fitting is to provide a measure of "goodness of fit" of the real time series to VAR (1) model, which we can use in the portfolio selection and trading parts of our overall algorithm.

3.1 Estimation of matrix A

We recall from our earlier discussion that we assume \mathbf{s}_t follows a stationary, first order autoregressive process as in equation (5). We first observe that if the number of assets n is greater than or equal to the length of the observed time series, then \mathbf{A} can be estimated by simply solving the linear system of equations:

$$\hat{\mathbf{A}}\mathbf{s}_{t-1} = \mathbf{s}_t. \quad (13)$$

This gives a perfect VAR(1) fit for our time series for cases where we have a large portfolio of potential assets (e.g. considering all 500 stocks which make up the S&P 500 index), from which a sparse mean-reverting subportfolio is to be chosen.

In most practical applications, however, the length of the available historical time series is greater than the number of assets considered, so \mathbf{A} is estimated using, for example, least squares estimation techniques, as in

$$\hat{\mathbf{A}} : \min_{\mathbf{A}} \sum_{t=2}^T \|\mathbf{s}_t - \mathbf{A}\mathbf{s}_{t-1}\|^2 \quad (14)$$

where $\|\cdot\|^2$ denotes the Euclidian norm.

Equating to zero the partial derivatives of the above expression with respect to each element of the matrix \mathbf{A} , we obtain the following system of equations:

$$\sum_{k=1}^n \hat{\mathbf{A}}_{i,k} \sum_{t=2}^T \mathbf{s}_{t-1,k} \mathbf{s}_{t-1,j} = \sum_{t=2}^T \mathbf{s}_{t,i} \mathbf{s}_{t-1,j} \quad \forall i, j = 1, \dots, n \quad (15)$$

Solving for $\hat{\mathbf{A}}$ and switching back to vector notation for \mathbf{s} , we obtain

$$\hat{\mathbf{A}} = \sum_{t=2}^T (\mathbf{s}_{t-1}^T \mathbf{s}_{t-1})^+ (\mathbf{s}_{t-1}^T \mathbf{s}_t), \quad (16)$$

where \mathbf{M}^+ denotes the Moore-Penrose pseudoinverse of matrix \mathbf{M} . Note that the Moore-Penrose pseudoinverse is preferred to regular matrix inversion, in order to avoid problems which may arise due to potential singularity of $\mathbf{s}_{t-1}^T \mathbf{s}_{t-1}$.

3.2 Estimation of the covariance matrix of W

Assuming that the noise terms in equation (5) are i.i.d. with $\mathbf{W}_t \sim N(0, \sigma I)$ for some $\sigma > 0$, we obtain the following estimate for σ using $\hat{\mathbf{A}}$ from (16) :

$$\hat{\sigma} = \sqrt{\frac{1}{n(T-1)} \sum_{t=2}^T \|\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1}\|^2}. \quad (17)$$

In the more general case that the terms of \mathbf{W}_t are correlated, we can estimate the covariance matrix \mathbf{K} , of the noise as follows:

$$\hat{\mathbf{K}} = \frac{1}{K-1} \sum_{t=2}^T (\mathbf{s}_t - \mathbf{A}\mathbf{s}_{t-1}) T (\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1}) \quad (18)$$

This noise covariance estimate will be used below in the estimation of the covariance matrix.

3.3 Estimation of covariance matrix G

There are two independent approaches to estimating the covariance of a VAR(1) process based on a sample time series. On the one hand, the sample covariance and various maximum likelihood-based regularizations thereof can provide a simple estimate and have been studied extensively for the more general case of multivariate Gaussian distributions [1, 4, 6, 14]. In our treatment, we take the approach of using the sample covariance matrix directly without any regularization or finding structure via maximum likelihood, as sparsifying and structure finding will be left for the more sophisticated apparatus of the sparse portfolio selection, explained in Section 2. As such, we will define $\hat{\mathbf{G}}_1$ as the sample covariance defined as

$$\hat{\mathbf{G}}_1 := \frac{1}{T-1} \sum_{t=1}^T (\mathbf{s}_t - \bar{\mathbf{s}})^T (\mathbf{s}_t - \bar{\mathbf{s}}), \quad (19)$$

where $\bar{\mathbf{s}}$ is the sample mean vector of the assets defined as

$$\bar{\mathbf{s}} := \frac{1}{T} \sum_{t=1}^T \mathbf{s}_t \quad (20)$$

On the other hand, starting from the description of the VAR(1) process in (5) and assuming the more general case that the terms of \mathbf{W}_t are correlated with covariance matrix \mathbf{K} , we must have

$$\mathbf{G}_t = \mathbf{A}^T \mathbf{G}_{t-1} \mathbf{A} + \mathbf{K}, \quad (21)$$

which implies that in the stationary case

$$\mathbf{G} = \mathbf{A}^T \mathbf{G} \mathbf{A} + \mathbf{K}. \quad (22)$$

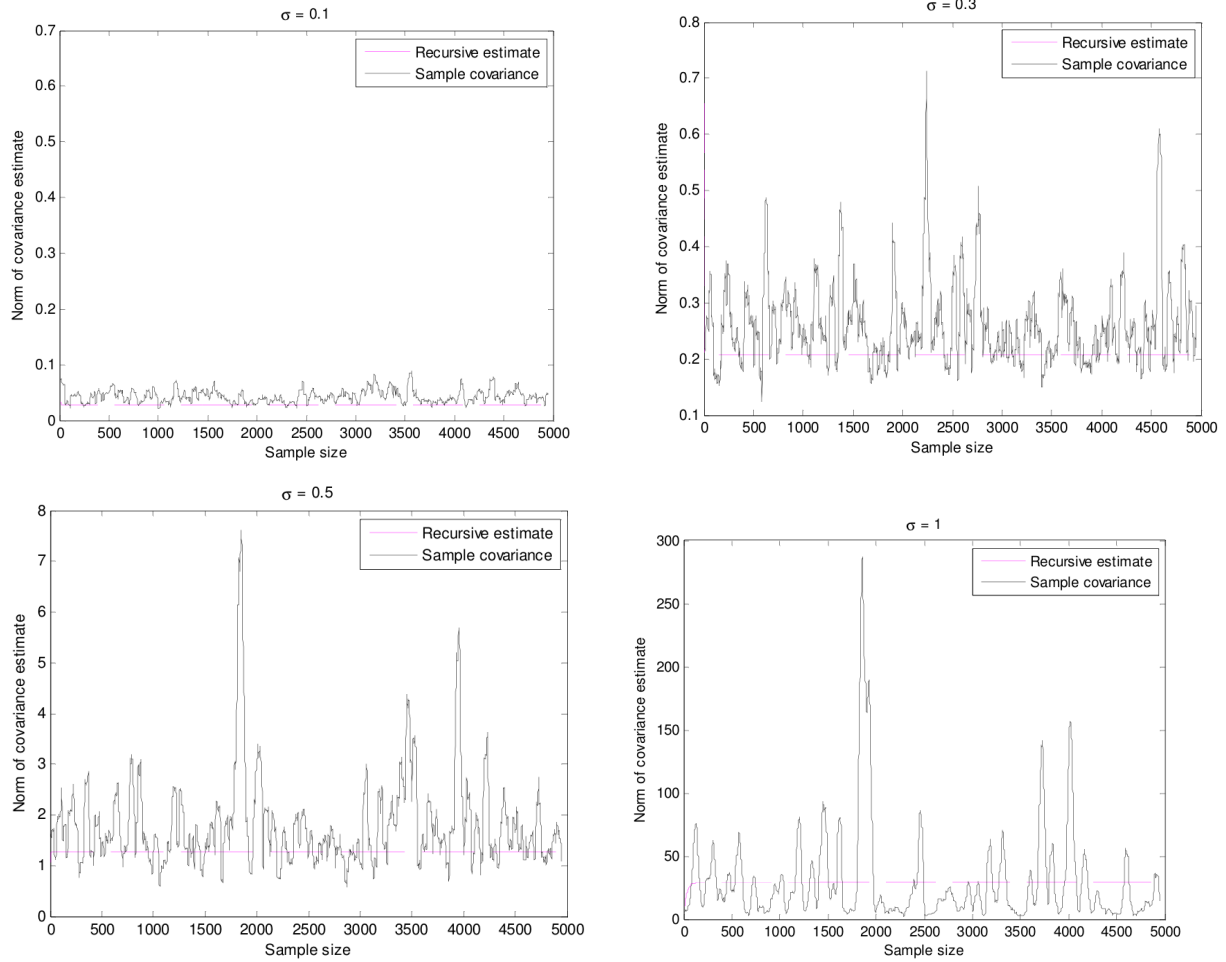


Fig. 3. Sample covariance and recursive covariance estimates over sliding windows of size 50 over 5000 samples for $\sigma = 0.1, 0.3, 0.5, 1$ (note the differences in scaling of the plots)

Having estimated \mathbf{A} and \mathbf{K} , as in the previous sections, this is a Lyapunov equation with unknown \mathbf{G} and can be solved analytically to obtain an independent covariance estimate $\hat{\mathbf{G}}_2$. One potential issue with this approach is that $\hat{\mathbf{G}}_2$ is not necessarily positive definite and, as such, it may not be a permissible covariance estimate. In order to overcome this issue, in case the solution of the Lyapunov equation is non-positive-definite, the following iterative numerical method can be used to obtain a permissible covariance estimate $\hat{\mathbf{G}}_2$:

$$\mathbf{G}(k+1) = \mathbf{G}(k) - \delta(\mathbf{G}(k) - \mathbf{A}^T \mathbf{G}(k) \mathbf{A} - \mathbf{K}), \quad (23)$$

where δ is a constant between 0 and 1, $\mathbf{G}(i)$ is the covariance matrix estimate on iteration i . Provided that the starting point for the numerical method, $\mathbf{G}(0)$, is positive definite (eg. the sample covariance matrix) and since our estimate of \mathbf{K} is positive definite, by construction, this iterative method will produce an estimate which will be positive definite. It can also be seen that with appropriate choice of δ and stopping condition, this numerical estimate will converge to the solution of the Lyapunov equation in (22), in case that is positive definite.

In Section 4, some numerical results are presented which show that for generated VAR(1) data, these two covariance estimates are equivalent, provided that appropriately sized sample

data is available for the given level of noise. However, for historical financial data, the two estimates can vary significantly. A large difference between the two estimates indicates a poor fit of the data to the VAR(1) model, hence we can define the following measure of model fit:

$$\beta := \|\hat{\mathbf{G}}_1 - \hat{\mathbf{G}}_2\|, \quad (24)$$

where $\|\mathbf{M}\|$ denotes the largest singular value of matrix \mathbf{M} .

4 Performance analysis

In this section, we will review some results of the numerical tests which were produced for validating the methods outlined earlier. We first tested the model parameter estimation methods on generated data to show their viability and observe their limitations. We then compared the effectiveness of the greedy search to the theoretically optimal exhaustive method on generated time series.

4.1 Performance of parameter estimation

In order to test our estimate $\hat{\mathbf{A}}$ as defined in (16), we generated $n=8$ -dimensional data with random \mathbf{A} matrix using equation (5) with i.i.d. noise $\mathbf{W}_t \sim N(0, \sigma I)$. Note that for stability of the resulting \mathbf{s}_t generated time series, we ensured that all eigenvalues

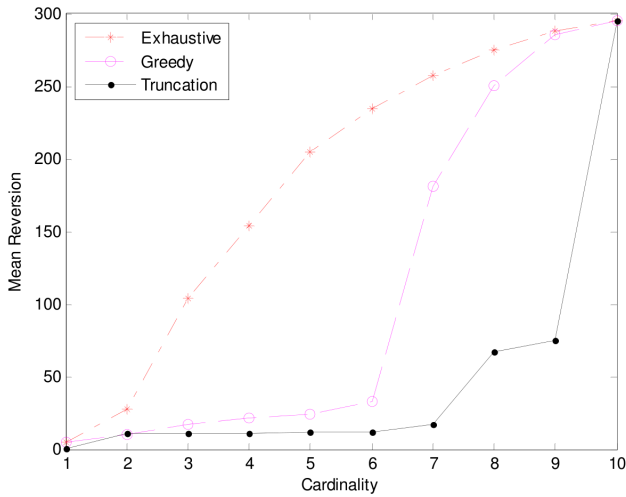


Fig. 4. Comparison of portfolio selection methods of various cardinality on 10-dimensional generated VAR(1) data.

of \mathbf{A} were smaller than 1. Then, for increasing sizes of t , we ran a number of independent tests and looked at the average error in norm. Figure 1 shows that our estimate gets progressively better with the size of the available time series and that it is remarkably resilient to increasing values of σ .

In a similar fashion, we compared the covariance estimate $\hat{\mathbf{G}}_1$ as defined in (19) to $\hat{\mathbf{G}}_2$, the solution of the Lyapunov equation, substituting $\hat{\mathbf{A}}$ and σ as estimated from the data into equation (22). We then considered the proximity of these two estimates on increasing sample sizes and increasing values of σ , taking the average of a number of independent tests. Figure 2 shows the results.

We see that there is good agreement between the two estimates of covariance even for relatively small amounts of data (100-300 data points) which gets increasingly better with the growth of the sample size. Note also that there is significant difference between the different amounts of noise on the process, smaller noise implies a better agreement between the two covariance indicators.

In our next numerical test, we used fixed sized sliding windows of various sizes over the generated VAR(1) sequence and we compared the sample covariance in the current time window with the current value of the recursive relation in (21). Figure 3 shows the values of the two estimates for varying levels of σ . We can see that both measures converge to the same level, but the covariance estimate obtained from (21) is more resistant to noise in the sample data.

4.2 Performance of portfolio selection and trading

In order to compare the portfolio selection methods outlined in Section 2, we again generated VAR(1) data with random \mathbf{A} matrix and noise with covariance matrix \mathbf{K} . We then used the methods of Section 3 to compute the estimates $\hat{\mathbf{A}}$, $\hat{\mathbf{K}}$ and $\hat{\mathbf{G}}$ and computed optimal sparse portfolios, maximizing the mean-reversion coefficient λ . We found that in a large number of

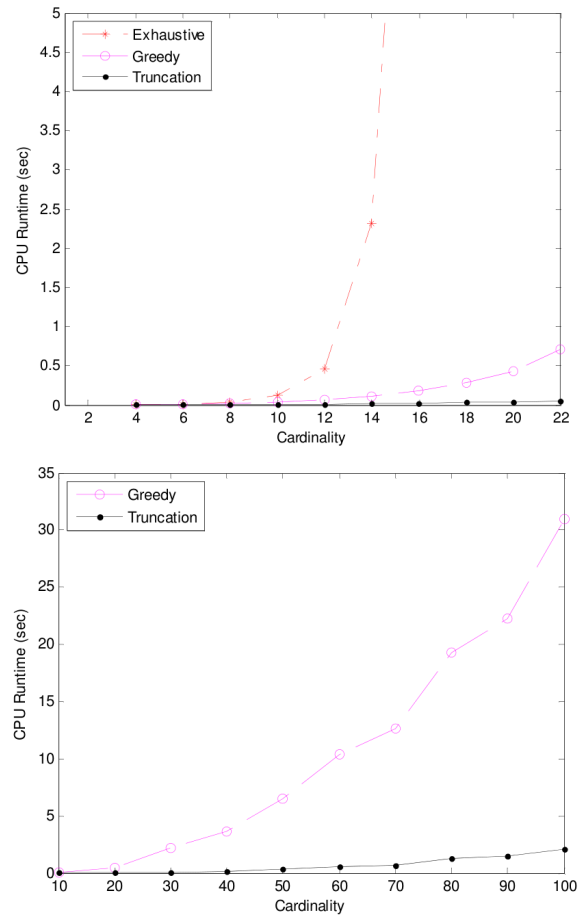


Fig. 5. CPU runtime (in seconds) versus total number of assets n , to compute a full set of sparse portfolios, with cardinality ranging from 1 to n , using the different algorithms.

cases, the greedy method yields portfolios whose mean reversion is close to the theoretical best, produced by the Exhaustive method. Having run 1000 simulations on independently generated VAR(1) data, we observed that the exhaustive method produced better mean reversion than the greedy method in 59.3% of the cases and outperformed the truncation method in 99.8% of the cases. The greedy method produced lambdas which were, on average, 2.26% worse than the optimal lambda found by the exhaustive search while truncation produced lambdas which were 7.34% worse on average. We also found a number of examples where greedy yielded significantly suboptimal portfolios where other polynomial time heuristic methods could be found to improve upon this. (for an example, see Figure 4).

In order to examine the runtime of the portfolio selection algorithms, we ran repeated simulations of selecting sparse portfolios from n assets for all cardinalities from 1 to n and plotting the total CPU time taken against n for each of the proposed methods (Figure 5). We observe that the truncation method is the fastest, taking less than 3 seconds on a Pentium 4, 3.80 GHz machine to select all 100 subportfolios of 100 assets. The same took over 30 seconds for the greedy method which suggests that while the truncation method is well suited for real-time algorithmic trading where sub-second algorithms are required for a

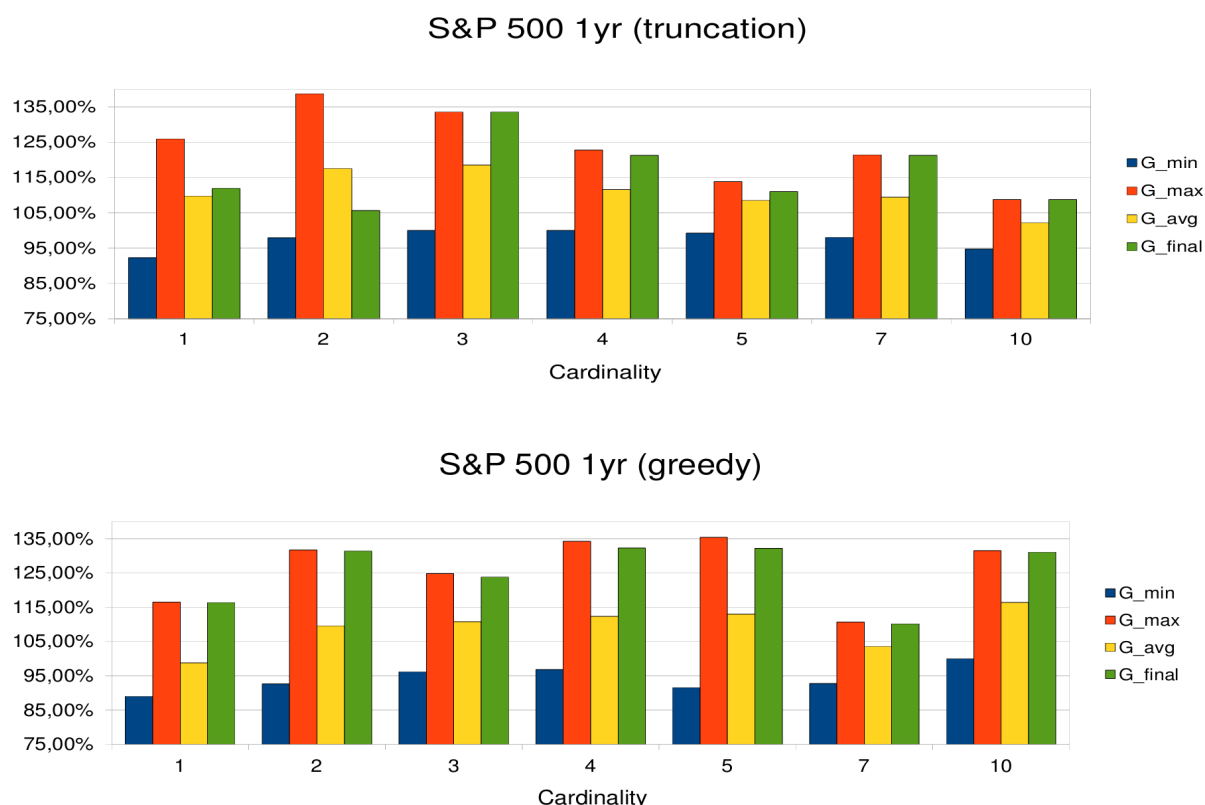


Fig. 6. Comparison of minimum return (G_{\min}), maximum return (G_{\max}), average return (G_{avg}) and final return (G_{final}) on S&P500 historical

data of the truncation and greedy methods for sparse mean-reverting portfolios of size 1, 2, 3, 4, 5, 7 and 10.

given cardinality, the greedy method could also be used for intraday electronic trading. The exhaustive search could also be viable for intraday trading for asset populations of 20 or under, but the run times become hours on our test hardware beyond 22 assets.

Finally, in order to test the economic viability of the methodology, we developed a simple converge trading methodology on historical time series, following the methodology of [5]. We consider daily close prices of the 500 stocks which make up the S&P500 index from July 2009 until July 2010. We use the methods of section 3 to estimate the model parameters on a sliding window of 8 observations and select sparse, mean reverting portfolios using the algorithms of Section 2. Considering portfolios of sparseness 3 and 4, the methods produced annual returns in the range of 23–34% (note that the return on the S&P500 index was 11.6% for this reference period). Detailed results are presented on Figure 6.

5 Conclusions and directions for future research

In this paper, we have examined the problem of selecting optimal sparse mean-reverting portfolios based on observed or generated time series. We have suggested new, more exact approaches for estimating the parameters \mathbf{A} , \mathbf{G} and \mathbf{K} and introduced a measure, β of goodness of fit to the model. We also introduced and analyzed the performance of the exhaustive portfolio selection method and the truncation based heuristic method as compared to the greedy method suggested by earlier research.

We have shown that the exhaustive method can be a viable practical alternative for smaller number of assets and it significantly outperforms the greedy method in some cases. We have also shown that the truncation method is significantly faster than the greedy method and therefore could be applied for real-time algorithmic trading where sub-second response times are required. We have demonstrated the economic viability of these methods by showing excess returns on historical daily close prices of stocks making up the S&P500 index.

Given the demonstrated relative suboptimality of the greedy heuristic, a fertile area of future research is the identification of other polynomial time heuristic algorithms which could improve upon the greedy solution. Furthermore, the practical viability of these heuristics could be further shown by constructing more complex trading methodologies and backtesting on a larger set of historical financial time series.

References

- 1 Banerjee O, El Ghaoui L, d'Aspremont A., *Model Selection Through Sparse Maximum Likelihood Estimation*, Journal of Machine Learning Research, **9**, (2008), 485–516.
- 2 Boguslavsky M, Boguslavskaya E, *Arbitrage Under Power*, Risk Magazine, (2004), 69–73.
- 3 Box GE, Tiao GC, *A canonical analysis of multiple time series*, Biometrika, **64**(2), (1977), 355–365.
- 4 D'Aspremont A, Banerjee O, El Ghaoui L, *First-Order Methods for Sparse Covariance Selection*, SIAM Journal on Matrix Analysis and its Applications, **30**(1), (2008), 56–66.

- 5 **D'Aspremont A**, *Identifying small mean-reverting portfolios*, Quantitative Finance, **11**(3), (2011), 351–364.
- 6 **Dempster A**, *Covariance selection*, Biometrics, **28**, (1972), 157–175.
- 7 **Dixit AK, Pindyck RS**, *Investment Under Uncertainty*, Princeton University Press; Princeton, NJ, 1994.
- 8 **Fama E, French K**, *Permanent and Temporary Components of Stock Prices*, The Journal of Political Economy, **96**(2), (1988), 246–273.
- 9 **Ito K**, *Stochastic Integral*, Proc. Imperial Acad., **20**, (1944), 519–524.
- 10 **Manzan S**, *Nonlinear Mean Reversion in Stock Prices*, Quantitative and Qualitative Analysis in Social Sciences, **1**(3), (2007), 1–20.
- 11 **Natarjan BK**, *Sparse approximate solutions to linear systems*, SIAM J. Comput., **24**(2), (1995), 227–234.
- 12 **Ornstein LS, Uhlenbeck GE**, *On the Theory of the Brownian Motion*, Physical Review, **36**(5), (1930), 823–.
- 13 **Poterba JM, Summers LH**, *Mean reversion in stock prices: Evidence and implications*, Journal of Financial Economics, **22**(1), (1988), 27–59.
- 14 **Rothman A, Bickel P, Levina E, Zhu J**, *Sparse permutation invariant covariance estimation*, Electronic Journal of Statistics, **2**, (2008), 494–515.